Standardized benchmark for field-level cosmological inference from galaxy surveys

Hugo Simonfroy[1], Arnaud de Mattia[1], François Lanusse[2]

[1]Université Paris-Saclay, CEA, IRFU, 91191, Gif-sur-Yvette, France
[2]Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France

Abstract

- Standard analyses of galaxy surveys rely on analytical models for the 2-pt statistics of the galaxy density field. Yet, gravity-driven structure formation is non-linear, introducing non-Gaussianity in the observed small-scale galaxy density field. This makes the standard 2-pt compression lossy on small scales, which are also more challenging to model analytically.
- To leverage as most as possible the information available in surveys data, we aim to perform field-level inference, i.e. we want to **directly estimate cosmological parameters from the observed galaxy density field, while reconstructing the initial conditions of the Universe.** High-fidelity cosmological simulations therefore becomes imperative to exploit information that can hardly be modeled analytically.
- A range of Markov Chain Monte Carlo (MCMC) methods have been developed to perform Bayesian inference over such simulations. How these methods compare, and which ones are more likely to scale to the size of Stage-IV galaxy surveys like DESI and Euclid is still currently unclear. To provide some answers and facilitate further research in sampling strategies, we introduce a standardized benchmark for field-level inference methods.

Differentiable model of galaxy density field

- Our model is composed of several blocks modeling different aspects of the cosmological evolution of the universe, as illustrated in Figure 1.
  1. Sample from a prior on cosmology Ω, initial linear field δ_L, and biases b.
  2. Initialize matter particles according to the initial field.
  3. Advect particles by a LPT+PM (Lagrangian Perturbation Theory + Particle Mesh) displacement to simulate Large Scale Structure (LSS) formation
  4. Apply Lagrangian bias to account for galaxy formation.
  5. Advect particles to account for contribution of galaxy peculiar velocities to redshift measurement, alias Redshift Space Distortion (RSD).
  6. Add observation noise to obtain the simulated observed galaxy density field δ_g.

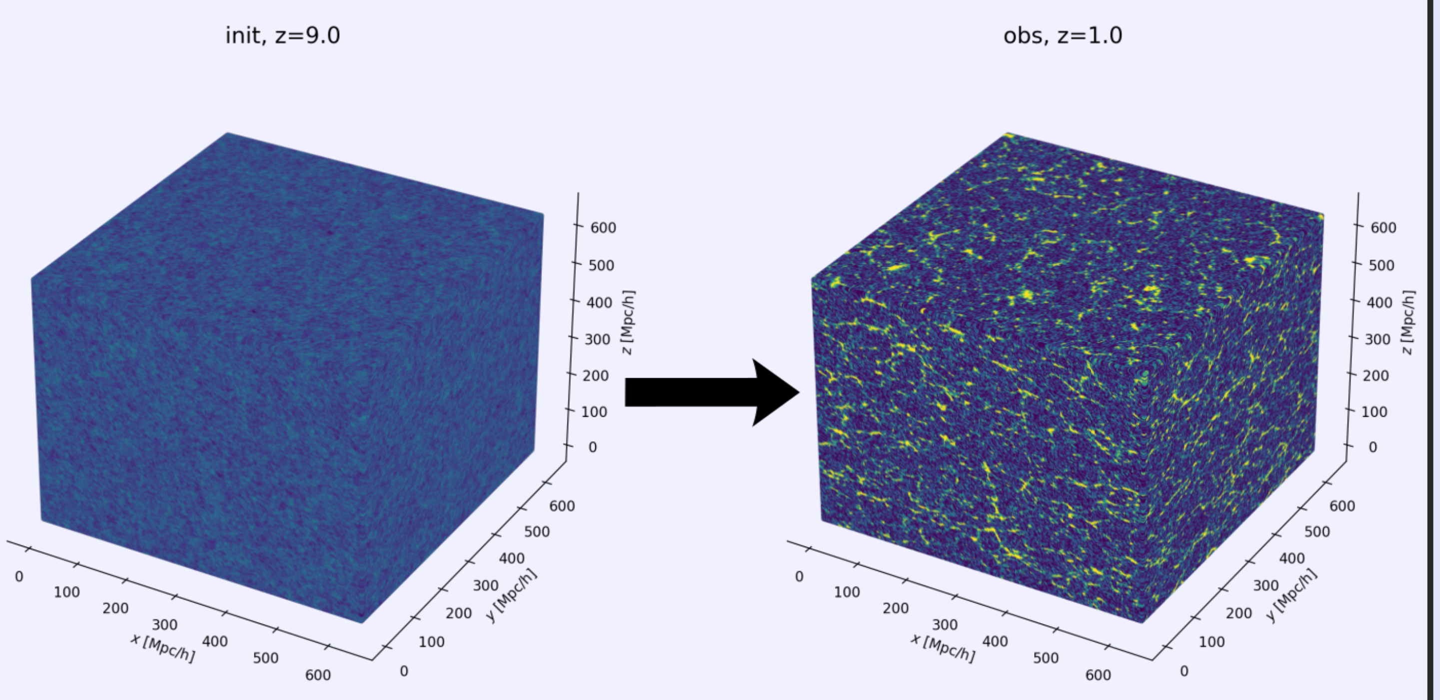

Figure 1: Model illustration. A 256^3 mesh, $(640\ \mathrm{Mpc/h})^3$ physical box, is evolved from initial redshift $z = 9$ to observed $z = 1$.

- In practice, we extensively use `JAX`, a `Python` library allowing GPU acceleration, Just-In-Time (JIT) compilation, Automatic vectorization/parallelization, and differentiation.
- We implement our model via `NumPyro`, a Probabilistic Programming Language (PPL) powered by `JAX`, which allows easy sampling, tractability and differentiation of our model. And to compute LPT+PM N-body displacement, we rely on `JaxPM` package.
- **Our model is formally defined by its joint probability $p(\Omega, \delta_L, b, \delta_g)$, and we aim to estimate full posterior $p(\Omega, \delta_L, b \mid \delta_g)$.** Cosmological constraints are then obtained from marginalization $p(\Omega \mid \delta_g) = \int p(\Omega, \delta_L, b \mid \delta_g) \mathrm{d}b\, \mathrm{d}\delta_L$.

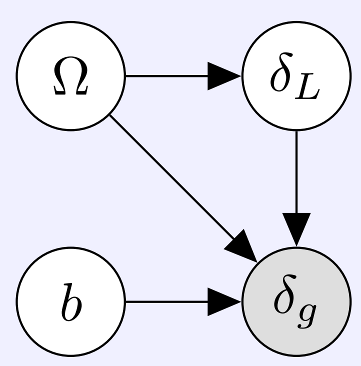

Figure 2: Model Bayesian network. White and grey nodes are latent and observed random variables respectively.

Benchmark metrics

- The Effective Sample Size (ESS) of a (possibly correlated) sample sequence is defined as the equivalent number of i.i.d. samples that yield the same statistical power. For a sample sequence of size N and autocorrelation ρ, ESS is then

$$N_{\text{eff}} := \frac{N}{\sum_{t=-\infty}^{+\infty} \rho_t} = \frac{N}{1 + 2\sum_{t=1}^{+\infty} \rho_t}$$

- We are specifically interested in uncertainty quantification, and therefore some specific posterior statistics, e.g. mean μ and standard deviation (std) σ. For a MCMC run with c chains, we can compute a (Normalized) Mean Square Error (NMSE) of any d-variate statistics θ of the samples

$$\mathrm{NMSE}(\theta) := \frac{1}{c^2 d} \sum_{d,c} \varepsilon^2_{d,c}(\theta) \qquad \text{where in particular } \varepsilon(\theta) := \begin{cases} (\hat{\mu} - \mu)/\sigma & \text{if } \theta \text{ is } \mu \\ \sqrt{2}(\hat{\sigma} - \sigma)/\sigma & \text{if } \theta \text{ is } \sigma \end{cases}$$

NMSE(θ) specifically probes the statistical power yielded by the sample sequence for the statistic θ. Normalization permits commensurability among parameters and with $1/N_{\text{eff}}$.

- Model evaluation is our main limiting computational factor for cosmological inference. We therefore **characterize the efficiency of a MCMC sampler by the number of model evaluations required to yield one effective sample $N_{\text{eval}}/N_{\text{eff}}$, and commensurably to yield unit error $N_{\text{eval}} \times \mathrm{NMSE}(\theta)$.**

Gradient-based MCMC methods

- Markov Chain Monte Carlo (MCMC) are sampling methods that rely on particle random walks (hence Markov Chain) to produce samples estimating a target density (hence Monte Carlo). It can be done by particles "time-average" alias Metropolis-Hastings (MH) type, or "space-average" alias particle filter type.
- We care about model differentiability because **state-of-the-art MCMC methods allow high-dimensional sampling by relying on the gradient of the model log-probability**. A typical gradient-based MCMC is Hamiltonian Monte Carlo (HMC) (Betancourt, 2018).
- To sample from density $p(q)$, HMC augments the sampling variables q by momentum variables p. The target $p(q)$ becomes $p(q,p) := e^{-\mathcal{H}(q,p)}$, where we introduce a mass matrix M and Hamiltonian

$$\mathcal{H}(q,p) := -\log p(q) + \frac{1}{2} p^\top M^{-1} p$$

As illustrated in the figure below (left), momentum p is resampled at each proposal step, then (p,q) follows the Hamiltonian dynamic during time length L and finishes with a new MH proposal.

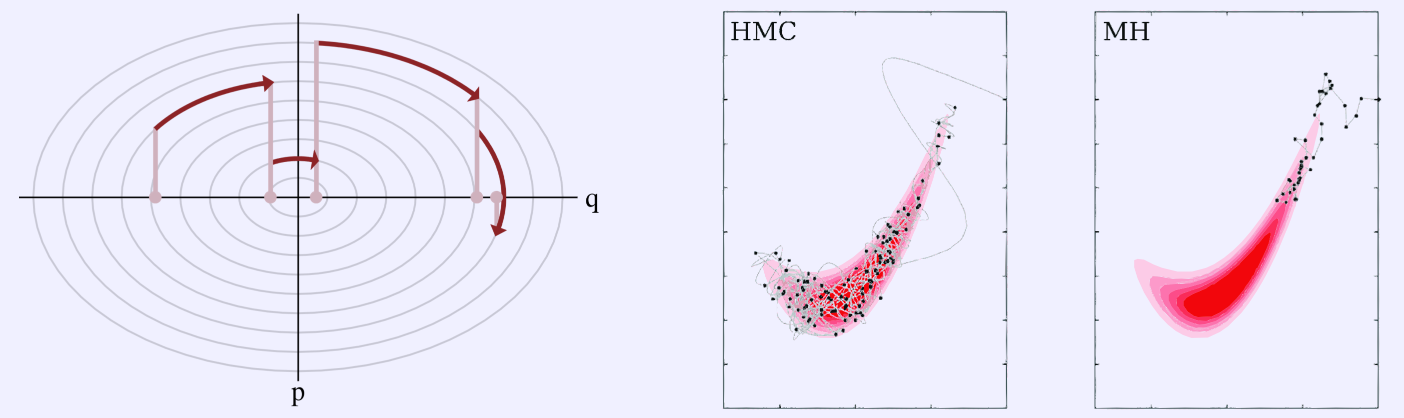

Figure 3: (left) HMC illustration in phase space. (right) Hamiltonian dynamics allows faster exploration of the density landscape therefore yielding less correlated samples than classical MH.

- Variations around HMC have also been proposed. The No U-Turn Sampler (NUTS) has trajectory length L automatically tuned, and samples can be drawn along the Hamiltonian trajectory, not only at the arrival.
- HMC can also be implemented within a Gibbs sampler, i.e. alternating sampling over parameter subsets. We test hereby a NUTS within Gibbs version, that we denote NUTSGibbs.

Experimental setup and results

- We consider a model setting with 64^3 mesh, $(640\ \mathrm{Mpc/h})^3$ box, 1LPT displacement, second order Lagrangian bias expansion (Modi, Chen, and White, 2020), RSD and Gaussian observational noise. A fiducial observation is generated from a Planck18-like cosmology prior, on which we condition our model.
- Our parameter space is composed of initial field δ_L, cosmology $\Omega = \{\Omega_m, \sigma_8\}$, and biases $b = \{b_1, b_2, b_{s^2}, b_{\nabla^2}\}$ for a total of $64^3 + 2 + 4$ parameters.
- For each tested sampler, we run 8 chains initialized on prior samples. For each sampler, step size $\mathrm{d}t$ and mass matrix M are tuned during a warmup phase, which ensure proper acceptance rate and conditioning. For NUTSGibbs, we divide sampling between initial field δ_L and the rest (Ω and b).
- We run each sampler for $N_{\text{eval}} \simeq 3 \times 10^7$ evaluations, and ensure posterior samples are consistent between each other, i.e. that samplers yield the same posterior statistics such as mean, std, and highest density regions. To compute NMSE, true posterior mean μ and std σ are estimated with all samples combined.

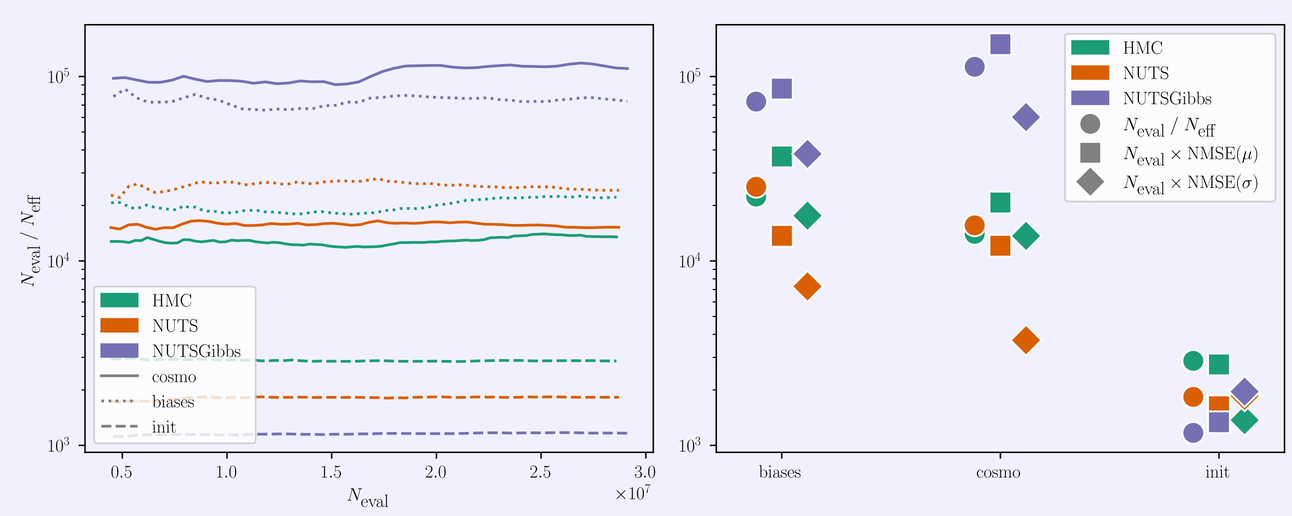

Figure 4: Benchmark results. (left) ESS metric along evaluations. (right) Comparison of the 3 metrics.

- We furthermore check that metrics $N_{\text{eval}}/N_{\text{eff}}$, and $N_{\text{eval}} \times \mathrm{NMSE}(\theta)$ are consistent with the number of evaluations (Figure 4, left), and consistent with each other (Figure 4, right).
- Results suggest **similar efficiency between HMC and NUTS samplers, and no particular advantage to splitting sampling between initial conditions and the rest of the parameters.**

Conclusion and prospects

- We introduced a standardized framework to benchmark field-level inference methods and test the impact of different settings with respect to consistent metrics. We compared 3 different MCMC samplers on a relatively small ($\simeq 64^3$) field-level inference problem.
- In the following, we plan to include more proposed MCMC methods into this benchmark, such as Micro Canonical Langevin Monte Carlo (MCLMC) (Bayer, Seljak, and Modi, 2023), and better tuned versions of within-Gibbs samplers closer to `BORG` framework (Lavaux, Jasche, and Leclercq, 2019). We also plan to test the impact of different model settings such as mesh dimension, resolution, and nuisance effects.

References

Bayer, Adrian E., Uros Seljak, and Chirag Modi (July 2023). "Field-Level Inference with Microcanonical Langevin Monte Carlo". en. In: arXiv:2307.09504. arXiv:2307.09504 [astro-ph, physics:physics, stat]. URL: `http://arxiv.org/abs/2307.09504`.

Betancourt, Michael (July 2018). "A Conceptual Introduction to Hamiltonian Monte Carlo". en. In: arXiv:1701.02434. arXiv:1701.02434 [stat]. URL: `http://arxiv.org/abs/1701.02434`.

Lavaux, Guilhem, Jens Jasche, and Florent Leclercq (Sept. 2019). "Systematic-free inference of the cosmic matter density field from SDSS3-BOSS data". en. In: arXiv:1909.06396. arXiv:1909.06396 [astro-ph]. URL: `http://arxiv.org/abs/1909.06396`.

Modi, Chirag, Shi-Fan Chen, and Martin White (Mar. 2020). "Simulations and symmetries". en. In: *Monthly Notices of the Royal Astronomical Society* 492.4. arXiv:1910.07097 [astro-ph], 5754–5763. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/staa251`.